

WHAT IS CLAIMED IS:

1 1. A computer program product, stored on a machine-readable medium, comprising
2 instructions operable to cause a programmable processor to:

3 search a document for one or more unambiguous words, where unambiguous words
4 are words that do not contain an ambiguous typesetting placeholder;

5 automatically add the one or more unambiguous words to a dictionary;

6 search the document for one or more ambiguous words, where ambiguous words are
7 words that do contain an ambiguous typesetting placeholder; and to

8 use the dictionary to resolve the one or more ambiguous words by resolving the
9 ambiguous typesetting placeholders occurring in each ambiguous word.

1 2. The computer program product of claim 1, wherein the instruction to automatically
2 add one or more ambiguous words to a dictionary comprises instructions to add the one or
3 more ambiguous words to an initially empty dictionary.

1 3. The computer program product of claim 1, wherein the instruction to automatically
2 add one or more ambiguous words to a dictionary comprises instructions to add the one or
3 more ambiguous words to a dictionary containing one or more unambiguous words located in
4 one or more documents that have been previously processed by the computer program.

1 4. The computer program product of claim 1, wherein the instruction to use the
2 dictionary to resolve the ambiguous typesetting placeholders in each ambiguous word,
3 comprises instructions operable to cause a programmable processor to:

4 create a set of candidate solutions for the ambiguous word, wherein each candidate
5 solution comprises one or more character strings created by resolving the one or more
6 ambiguous typesetting placeholders in the ambiguous word, and wherein the set of candidate
7 solutions comprises all possible combinations of resolutions of the one or more typesetting
8 placeholders;

9 search the dictionary for the one or more character strings in each candidate solution;
10 and

11 use the dictionary search result to resolve the one or more ambiguous typesetting
12 placeholders in the ambiguous word.

1 5. The computer program product of claim 4, wherein the instruction to create a set of
2 candidate solutions for an ambiguous word having N binary-resolvable typesetting
3 placeholder ambiguities, comprises instructions to create a set of 2^N candidate solutions.

1 6. The computer program product of claim 4, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to resolve the one or more ambiguous
4 typesetting placeholders in conformity with the one or more resolutions used to create a
5 member of the set of candidate solutions when the dictionary search matches only that
6 member of the set of candidate solutions.

1 7. The computer program product of claim 4, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to prompt a user to manually resolve the one
4 or more ambiguous typesetting placeholders in the ambiguous word when the dictionary
5 search fails to match any member of the set of candidate solutions.

1 8. The computer program product of claim 4, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to prompt a user to manually resolve the one
4 or more ambiguous typesetting placeholders in the ambiguous word when the dictionary
5 search matches a plurality of members of the set of candidate solutions.

1 9. The computer program product of claim 4, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to resolve the one or more ambiguous
4 typesetting placeholders in conformity with the one or more resolutions used to create the
5 candidate solution having the largest word when the dictionary search matches a plurality of
6 members of the set of candidate solutions.

1 10. The computer program product of claim 9, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to resolve the one or more ambiguous
4 typesetting placeholders in conformity with the one or more resolutions used to create the
5 candidate solution having the fewest words when the dictionary search matches a plurality of
6 members of the set of candidate solutions.

1 11. The computer program product of claim 4, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to resolve the one or more ambiguous
4 typesetting placeholders in conformity with the one or more resolutions used to create the
5 candidate solution having the smallest word when the dictionary search matches a plurality
6 of members of the set of candidate solutions.

1 12. The computer program product of claim 11, wherein the instruction to use the
2 dictionary search result to resolve the one or more ambiguous typesetting placeholders in the
3 ambiguous word, further comprises instructions to resolve the one or more ambiguous
4 typesetting placeholders in conformity with the one or more resolutions used to create the
5 candidate solution having the most words when the dictionary search matches a plurality of
6 members of the set of candidate solutions.

1 13. The computer program product of claim 4, wherein the ambiguous typesetting
2 placeholders comprise hyphens resolvable as hard hyphens or soft hyphens.

1 14. The computer program product of claim 14, further comprising instructions operable
2 to cause a programmable processor to output the character code for the correct ambiguity
3 resolution.

1 15. The computer program product of claim 4, wherein the ambiguous typesetting
2 placeholders comprise white space between characters resolvable as blank space or kerning
3 space.

1 16. The computer program product of claim 16, further comprising instructions operable
2 to cause a programmable processor to add space to an ambiguous white space resolved to be
3 blank space and to remove space from an ambiguous white space resolved to be kerning
4 space.

1 17. A computer program product, stored on a machine-readable medium, comprising
2 instructions operable to cause a programmable processor to:

3 search the document for one or more unambiguous words, where unambiguous words
4 are words that do not contain an ambiguous typesetting placeholder;

5 automatically add the one or more unambiguous words to a dictionary;

6 search the document for an ambiguous word, where an ambiguous word is a word
7 that does contain an ambiguous typesetting placeholder;

8 create a set of candidate solutions for the ambiguous word, wherein each candidate
9 solution comprises one or more character strings created by resolving the one or more
10 ambiguous typesetting placeholders in the ambiguous word, and wherein the set of candidate
11 solutions comprises all possible combinations of resolutions of the one or more typesetting
12 placeholders;

13 search the dictionary for the one or more character strings in each candidate solution
14 of the ambiguous word;

15 resolve the one or more ambiguous typesetting placeholders in conformity with the
16 one or more resolutions used to create a member of the set of candidate solutions when the
17 dictionary search matches only that member of the set of candidate solutions;

18 prompt a user to manually resolve the one or more ambiguous typesetting
19 placeholders when the dictionary search fails to match any member of the set of candidate
20 solutions; and to

21 prompt a user to manually resolve the one or more ambiguous typesetting
22 placeholders when the dictionary search matches a plurality of members of the set of
23 candidate solutions.

1 18. A method for resolving an ambiguous word in an electronic document,
2 comprising:

3 searching the document for unambiguous words, where unambiguous words are
4 words that do not contain one or more ambiguous typesetting placeholders;
5 automatically adding the unambiguous words to a dictionary;
6 searching the document for an ambiguous word, where an ambiguous word is a word
7 that contains one or more ambiguous typesetting placeholders; and
8 using the dictionary to resolve the ambiguous word by resolving the one or more
9 ambiguous typesetting placeholders occurring in the word.

1 19. The method of claim 1, wherein the step of using the dictionary to resolve the
2 ambiguous word by resolving the one or more ambiguous typesetting placeholders, further
3 comprises:

4 creating a set of candidate solutions for each ambiguous word, wherein each
5 candidate solution comprises one or more character strings created by resolving the one or
6 more ambiguous typesetting placeholders in the ambiguous word, and wherein the set of
7 candidate solutions comprises all possible typesetting placeholder resolution combinations;

8 searching the dictionary for the one or more character strings in each candidate
9 solution; and

10 using the dictionary search to resolve the one or more ambiguous typesetting
11 placeholders in the ambiguous word.

1 20. The method of claim 20, further comprising resolving the one or more ambiguous
2 typesetting placeholders in conformity with the one or more resolutions used to create a
3 member of the set of candidate solutions when the dictionary search only matches that
4 member of the set of candidate solutions.

1 21. The method of claim 20, further comprising prompting a user to manually resolve the
2 one or more ambiguous typesetting placeholders when the dictionary search fails to match
3 any member of the set of candidate solutions.

1 22. The method of claim 20, further comprising prompting a user to manually resolve the
2 one or more ambiguous typesetting placeholders when the dictionary search matches a
3 plurality of members of the set of candidate solutions.

1 23. The method of claim 20, further comprising resolving the one or more ambiguous
2 typesetting placeholders in conformity with the one or more resolutions used to create the
3 candidate solution having the largest word when the dictionary search matches a plurality of
4 members of the set of candidate solutions.

1 24. The method of claim 20, further comprising resolving the one or more ambiguous
2 typesetting placeholders in conformity with the one or more resolutions used to create the
3 candidate solution having the smallest word when the dictionary search matches a plurality
4 of members of the set of candidate solutions.

1 25. The method of claim 20, wherein the ambiguous typesetting placeholders comprise
2 ambiguous hyphens resolvable into hard hyphens or soft hyphens, further comprising
3 outputting the character code for the correct ambiguity resolution.

1 26. The method of claim 20, wherein the ambiguous typesetting placeholder comprises an
2 ambiguous white space between characters resolvable to a blank space or a kerning space,
3 further comprising adding space to an ambiguous white space resolved to be blank space and
4 removing space from an ambiguous white space resolved to be kerning space.